OPTIMIZING TIMELY DIABETES DETECTION THROUGH ADVANCED PREDICTIVE MODELING AND PATIENT MEDICAL STATISTICS

Vineet Mehan

Professor, AIT CSE, Chandigarh University, Gharuan,
Mohali, Punjab 140413, India
mehanvineet@gmail.com
http://< cucet.cuchd.in>

Abstract

Diabetes has become prevalent across a significant portion of the global population. While medication offers one avenue for disease management, its dosage often escalates with age, eventually necessitating insulin injections. Alternatively, lifestyle adjustments provide another means of control. Detecting diabetes early holds the potential to enhance patient well-being and avert the onset of the disease. This study undertakes a thorough parametric analysis of diabetes, comparing Linear, Multi-Linear, Polynomial, and Logistic Regression Models. The goal is to identify the optimal approach for predicting diabetes. Leveraging a dataset encompassing 768 patients from the National Institute of Diabetes and Digestive and Kidney Diseases, USA, the research considers both Patient Medical Statistics (PMS) and a Combination of Patients Medical Statistics (CPMS), featuring variables such as Pregnancy, Blood Sugar (BS), Blood Pressure (BP), Body Fat (BF), Insulin Level (IL), Body Mass Index (BMI), Diabetes Pedigree Function (DPF), and Age. Evaluation metrics, including Root Mean Square Error (RMSE) and Coefficient of Determination (R2), guide model selection. By employing R2, the ranking of PMS and CPMS is determined to yield a highly efficient model. Empirical findings indicate the Multiple Linear Regression model as presenting the lowest RMSE among all regression models. However, the comparative assessment underscores the suitability of Logistic Regression due to its discrete nature, demonstrating superior accuracy in prediction compared to other models.

***Keywords*: Diabetes prevalence; Disease management; Early diabetes detection; Parametric analysis; Predictive modeling.**

1. Introduction

Diabetes is one of the most common disease that is prevalent in most of the families [1]. An excess of glucose level in the blood is observed in diabetes [2]. The disease is caused when the insulin secreted by pancreas is zero or not sufficient to allow the glucose to be absorbed by the cells of the body [3]. A few common symptoms of diabetes are: a person feels thirsty all the time; vision becomes a bit blurred; weight loss happens; there is an increase in urination; hunger is enhanced; skin starts becoming dry etc. [4].

Diabetes is generally categorized into two types: Type1 [5], [6] and Type 2 [7], [8] diabetes. Type1 diabetes happens when the body stops producing insulin [9]. Generally, this type of diabetes happens in childhood or in young adults although the disease can happen any time. It is often said that type1 diabetes is common among people who have genetic diabetes problem. People with type 1 diabetes are advised to take pills to fulfill the insulin need. In some cases injection is also taken by the diabetic person.

Type 2 diabetes happens when the body produces insulin, but it is not sufficient to fulfill the body requirements [10]. Generally, this type of diabetes happens in adults with age greater than 45, although the disease can happen any time. Type 2 diabetes happen due to age, wrong eating habits, being overweight etc. A healthy lifestyle is the best known prevention method for diabetes disease. Insulin pills and injections are also a part of type 2 diabetes treatment.

As per the World Health Organization report nearly 422 million people were diagnosed with diabetes in 2014 [11]. Since then this percentage is continuously rising. Death rate is also increasing due to diabetes. Diabetes can lead to heart attacks, nerve damages, blindness, kidney failure, prone to infectious diseases like COVID-19 [12]. Considering diabetes to be a chronic disease, the present research work is done in the field of diabetes so

as to provide an intelligent solution for early detection of diabetes. The paper is organized as follows: Literature Review is given in section 2; Regression Models for Diabetes Prediction are laid in section 3; Experimental Analysis is demonstrated in section 4; Conclusion is given in section 5; Finally, references are laid in section 6.

2. Literature Review

A system for controlling diabetes was proposed by Patek et. al. in 2023 by giving a three-layer architecture [13]. Insufficiency of insulin was projected as a major reason for Type-1 Diabetes. Within the layered architecture the amount of insulin requirement for a particular patient was identified. The entire model worked in real time where modular functions were recognized at each interface of the model. Wuttichai et. al. in 2012 identified various parameters which lead to detection of Diabetes [14]. The parameters included Body Mass Index, Weight Circumference, Blood Pressure, Age and Family History etc. Logistic Regression and Back Propagation Neural Network technique was compared in the research work. Accuracy and Root Mean Square Error parameters were considered in this work.

A model where the Type-1 Diabetes could be identified early in the patients was suggested by Zhao in 2023 [15]. The model focused on time and computational complexity for early detection of the disease. Difference between the old and the new subjects for diabetes identification were explored to find the essential gaps. The gaps were then explored for rapid identification of the disease. The model proved to be cheap in comparison with the old methods. Abbas et. al. in 2019 [16] proposed a system for identification of diabetes using a machine learning model. Healthy population was chosen in the research work. Dataset was taken from a study that was going on at San Antonio. Type-2 diabetes identification was done in this work. Support Vector Machine(SVM) was chosen as a model for diabetes prediction. Accuracy projected by the proposed system was 84.1%. Glucose parameter was considered to be of high probability in the development of diabetes.

A linear regression model for predicting the diabetes was given by Zhang et. al. in 2016 [17]. It was stated that relationships between variables could easily be identified by linear regression. Glycosylated hemoglobin parameter was used in the research work. Expectation maximization method was used in order to trace the missing data in the dataset. The model efficiency was evaluated for different missing rates. Zou et. al in 2020 [18] proposed a multi variable linear regression model for identification of the diabetes. Diabetes detection in early pregnancy was identified within this work. Apart from pregnancy other predictor variables were BMI, family history and circumference of the fetus. Least square method and Gradient decent parameters were identified for knowing the accuracy of the model.

Teja et. al. [19] identified a number of algorithms by which early diabetes can be predicted. Additional polynomial structures were used for detection of the disease. SVM, K Nearest Neighbors (KNN), Decision Tree, and Random forest were applied in the research work. The algorithms were compared in terms of accuracy. Venkatesh and Saravanan in 2022 [20] discussed a polynomial regression model. Dataset based on crops was taken in the work. Linear regression and polynomial regression techniques were compared in the work. Experimental results proved that the linear regression model was better than the polynomial regression.

Oraz and Luo [21] proposed a system based on multi factors for prediction of diabetes. In this work, features that were most important for identification of diabetes were recognized. Geographic distribution was also taken into account. Country level diabetes prevalence was also revealed in the work. The research gave a significant result were the policies can be framed for the prevention of the disease. Simone et. al. made [22] a custom model for identification of diabetes based on linear regression. Various regressions models were compared in the research work. Parameters for evaluation of models included Root Mean Square Error (RMSE) and Coefficient of Determination. It was evaluated that linear models proved to be beneficial that non-linear models.

Ganesh et. al. [23] proposed a technique of diabetes detection on the basis of logistic regression model. Type-2 diabetes was identified in the work. Indian dataset was taken for the research work. Feature normalization technique was also adopted in the work so as to scale down the feature in equal proportion. Liu Lei [24] proposed a model for diabetes identification using logistic regression model. The work also projected a linear regression model. Both the models were compared so as to find out the better accuracy results. For two features, logistic regression proved to give better results than the linear regression.

Mangal and Jain proposed a research work for evaluating the performance of the models that are used for identification of diabetes [25]. Random Forest Machine Learning (RFML) algorithm was adopted in the work. The research projected an accuracy of 99% using the proposed technique. In order to identify the diabetes disorder due to increased blood sugar level a research work was proposed by Faruque et. al [26]. Dataset taken for the research work was for adult population. Four different algorithms were compared in the research work

which were SVM, KNN, Decision Tress and Naïve' Bayes. The work projected Decision Trees to be a good algorithm that gave the highest efficiency.

3. Regression Models for Diabetes Prediction

Regression analysis model the relationship between independent (predictor) variables and dependent variable (target). It helps us to identify how the value of the dependent variable is changing corresponding to independent variables. Regression technique helps to find the correlation between variables on the basis of supervised learning. It enables to predict the continuous output variable based on one or more predictor variables. The paper explores various regression models for early diabetic prediction.

3.1. Linear Regression

Linear regression shows the linear relationship between the independent variables and the dependent variable [27]. It is used to model the association between two continuous variables. The objective is to predict the value of an output variable based on the value of an input variable. Pregnancy, Blood Sugar (BS), Blood Pressure (BP), Body Fat (BF), Insulin Level (IL), Body Mass Index (BMI), Diabetes Pedigree Function (DPF), Age will be considered independent variables. The final output i.e. Diabetes is there or not, will be the dependent variable. The mathematical equation for Linear regression is given in equation (1) [28] below

$$Y = aX + b \quad (1)$$

where, Y is the dependent variable (target variable), X is independent variables (predictor variables), a and b are the linear coefficients.

Key values observed from the linear regression include Slope of the regression line; Intercept of the regression line; Rvalue: Correlation coefficient; Pvalue: Probability-value for a hypothesis test; Standard error of the estimated gradient/slope. It is significant to know how the relationship between the independent variables and dependent variable is. If there is no relationship, the linear regression cannot be used to predict anything. This relationship - the coefficient of correlation(r) value ranges from -1 to 1, where 0 means no relationship, and 1 (and -1) means 100% related. Positive correlation happens if one parameter is increasing the other parameter is also increasing. Negative correlation happens if one parameter is increasing the other parameter is decreasing. Code for implementing linear regression is given in Fig. 1

```
import pandas
from sklearn import linear_model
from sklearn.model_selection import train_test_split
import numpy as np

#read the file
df = pandas.read_csv("3_diabetes_scaled.csv")

#take one independent variable
X= df.iloc[:, [0]].values
#take one dependent variable
y = df['Output']

#split between training and test set
x_train, x_test, y_train, y_test = train_test_split(X, y, test_size=.25, random_state=0)

#Apply linear regression to the dataset
regr = linear_model.LinearRegression()
regr.fit(x_train, y_train)

#Predict the output by passing the x_test variable.
y_pred = regr.predict(x_test)

#Test the accuracy of model. (Score is calculated using R2 metric by default)
regr.score(x_test,y_test)

# Root Mean Squared Error
Root_MSE = np.sqrt(np.square(np.subtract(y_test,y_pred)).mean())
print(Root_MSE)
```

Fig. 1. Code for implementing Linear Regression

3.2. Multiple Linear Regression (MLR)

Multiple linear regression exploits linear regression by using more than one independent variable. In MLR prediction is done based on two or more variables [29] [30]. Seven different coalitions are made for Patients Medical Statistics (PMS) which are shown in Table 1.

Table 1. Combination of Patients Medical Statistics (CPMS)

CPMS	Ordered PMS Statistics w.r.t. Linear Regression
C1	BS + Pregnancy
C2	BS + Pregnancy + BMI
C3	BS + Pregnancy + BMI + Age
C4	BS + Pregnancy + BMI + Age + DPF
C5	BS + Pregnancy + BMI + Age + DPF +BF
C6	BS + Pregnancy + BMI + Age + DPF +BF + BP
C7	BS + Pregnancy + BMI + Age + DPF +BF + BP + IL

Coding the variations of PMS i.e. CPMS using MLR is given in Fig. 2.

```
#Taking Combination of Variables
X1= df.iloc[:, [1,0]].values
X2= df.iloc[:, [1,0,5]].values
X3= df.iloc[:, [1,0,5,7]].values
X4= df.iloc[:, [1,0,5,7,6]].values
X5= df.iloc[:, [1,0,5,7,6,3]].values
X6= df.iloc[:, [1,0,5,7,6,3,2]].values
X7= df.iloc[:, [1,0,5,7,6,3,2,4]].values

#Taking one dependent variable
y = df['Output']
```

Fig. 2. Coding CPMS using MLR

3.3. Polynomial Regression (PR)

Polynomial regression identifies the non-linear relationship between independent variables and dependent variable [31]. A curve is obtained for the relationship instead of a straight line. Polynomial equation of degree two is used in our model for individual parameters as well as the CPMS. The mathematical equation for Polynomial regression is given in equation (2) [32] below

$$Y = aX^2 + bX + c \quad (2)$$

where, Y is the dependent variable (target variable), X is independent variables (predictor variables), a and b are the linear coefficients. Coding the polynomial features is given in Fig. 3. Polynomial regression of degree 2 is computed for PMS and CPMS.

```
poly_features =PolynomialFeatures(degree=2)
X_poly =poly_features.fit_transform(X)
```

Fig. 3. Coding the Polynomial Features

3.4. Logistic Regression (LR)

Logistic regression predicts the categorical dependent variable using a given set of independent variables [33]. In Logistic regression, instead of fitting a regression line, a "S" shaped logistic function is used to predict output.

The Sigmoid function uses the concept of the threshold value, which defines the probability of either 0 or 1 which is shown in Fig. 4 [34]. Values above the threshold reaches to 1, and a value below the threshold reaches to 0. Coding the logistic regression is given in Fig. 5.

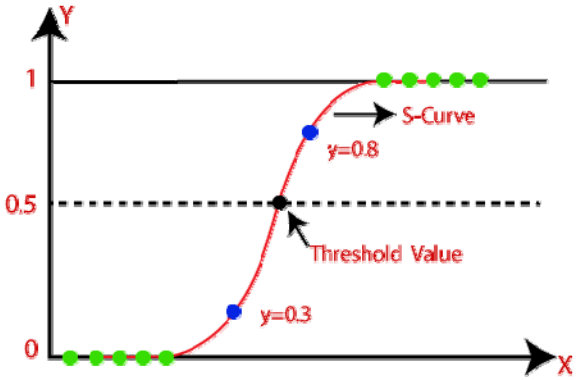


Fig. 4. Sigmoid Function

```
regr = linear_model.LogisticRegression()
regr.fit(x_train, y_train)
```

Fig. 5. Coding the Logistic Regression

4. Experimental Results

For the research work the diabetes dataset is collected from National Institute of Diabetes and Digestive and Kidney Diseases, USA [35]. 768 patient's data are identified from the dataset. Various parameters under which the data fall are Pregnancy, BS, BP, BF, IL, DPF, Age and Output. A total of 8 independent variables and 1 dependent variable are taken. Dataset of five patients out of 768 patients is depicted in Table 2.

Table 2. Diabetes Dataset for 5 Patients

Pregnancy	BS	BP	BF	IL	BMI	DPF	Age
6	148	72	35	0	33.6	0.627	50
1	85	66	29	0	26.6	0.351	31
8	183	64	0	0	23.3	0.672	32
1	89	66	23	94	28.1	0.167	21
0	137	40	35	168	43.1	2.288	33

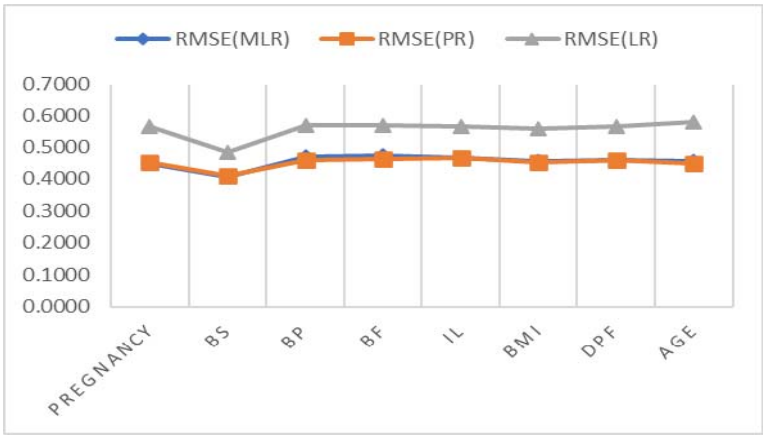
Feature scaling is applied to prevent overshadowing. Normalization of the dataset is done under feature scaling. Normalization ensures to make every data point fall under the same scale in order to ensure significant contribution of each feature. A scaled down version of the dataset for five patients out of 768 patients is depicted in Table 3.

Table 3. Scaled Diabetes Dataset for 5 Patients

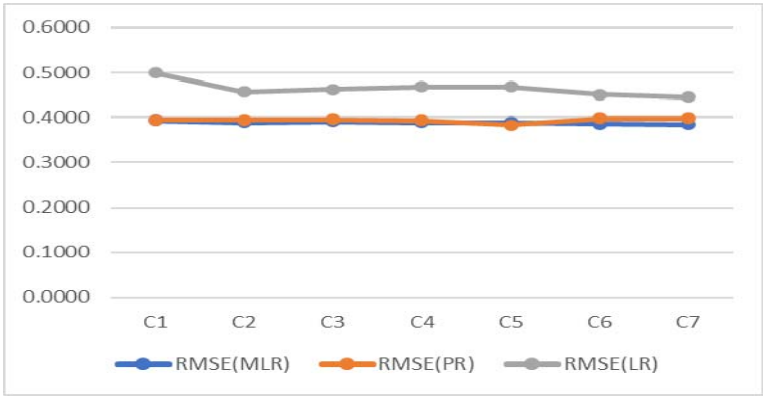
Pregnancy	BS	BP	BF	IL	BMI	DPF	Age
0.639947	0.848324	0.149641	0.907270	-0.692891	0.204013	0.468492	1.425995
-0.844885	-1.123396	-0.160546	0.530902	-0.692891	-0.684422	-0.365061	-0.190672
1.233880	1.943724	-0.263941	-1.288212	-0.692891	-1.103255	0.604397	-0.105584
-0.844885	-0.998208	-0.160546	0.154533	0.123302	-0.494043	-0.920763	-1.041549
-1.141852	0.504055	-1.504687	0.907270	0.765836	1.409746	5.484909	-0.020496

Root Mean Square Error (RMSE) is computed for PMS and CPMS using equation (3) [36]. In MLR when the number of features taken are one, then it is equivalent to linear regression. It is for this reason that no separate curve is made for linear regression. The graph shown in Fig. 6 depicts the average deviation between the definite and the predicted values. A lower value of RMSE depicts how well a model can be utilized for the prediction of diabetes.

$$RMSE = sqrt\left(\frac{\sum(Actual - Predictaion)^2}{Number\ of\ Observations}\right) \quad (3)$$



a) RMSE of PMS



b) RMSE of CPMS

Fig. 6. a) RMSE of PMS b) RMSE of CPMS

Accuracy of models is identified for each parameter as shown in Fig. 7. The entire PMS are computed for MLR, PR and LR models. R2 Score [37] identifies the accuracy of the models which is given in equation (4). Seeing the result, it is revealed that the best parameters for identifying the diabetes in an ordered arrangement are BS, Pregnancy, BMI, Age, DPF, BF, IL and BP.

$$R^2=1- \text{ RSS/TSS} \qquad (4)$$

where, RSS is Sum of squared residuals and TSS is Total sum of squares.

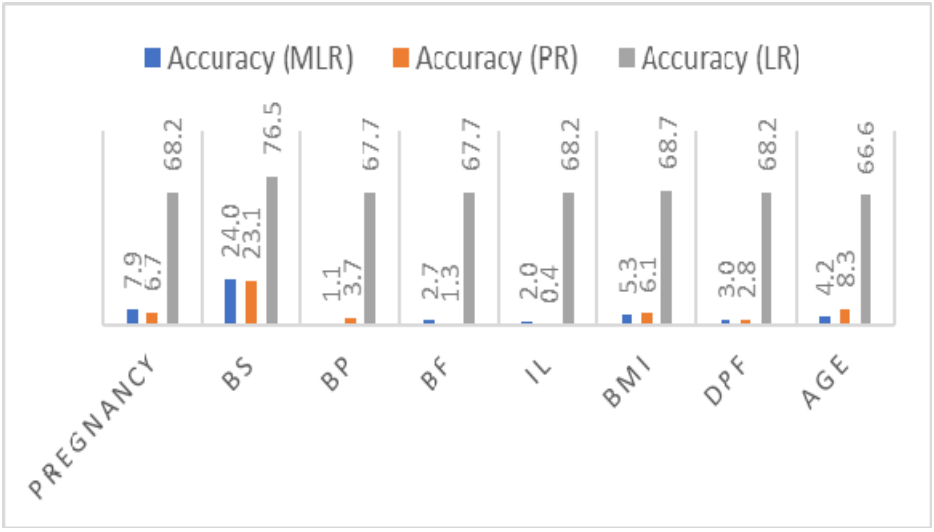


Fig. 7. Accuracy of PMS

A comparative analysis of MLR, PR and LR for each parameter in PMS is shown in Table 4. A significant percentage increase of accuracy is identified when Logistic Regression is used instead of Multiple Linear Regression and Polynomial Regression. The discrete nature of Logistic Regression is making a significant contribution in the enhancement of accuracy.

Table 4. Comparative Analysis of MLR, PR and LR for individual parameters

PMS	Accuracy (MLR)	Accuracy (PR)	Accuracy (LR)	Diff.(LR-MLR)	Diff.(LR-PR)	%Inc wrt MLR
Pregnancy	7.9	6.7	68.2	60.3	61.5	763.5
BS	24.0	23.1	76.5	52.5	53.4	218.8
BP	1.1	3.7	67.7	66.6	64.0	6054.5
BF	2.7	1.3	67.7	65.0	66.4	2407.4
IL	2.0	0.4	68.2	66.2	67.8	3310.0
BMI	5.3	6.1	68.7	63.4	62.6	1196.2
DPF	3.0	2.8	68.2	65.2	65.4	2173.3
Age	4.2	8.3	66.6	62.4	58.3	1485.7

Accuracy of the model is identified w.r.t. the ordered arrangement of parameters in Fig. 8. The parameters are combined together stepwise, so as to see if the accuracy of the model is enhanced or not. The entire CPMS are computed for the MLR, PR and Logistic Regression. R2 score is used to identify the accuracy of the model. Seeing the result, it is revealed that on uniting the parameters the accuracy is enhanced.

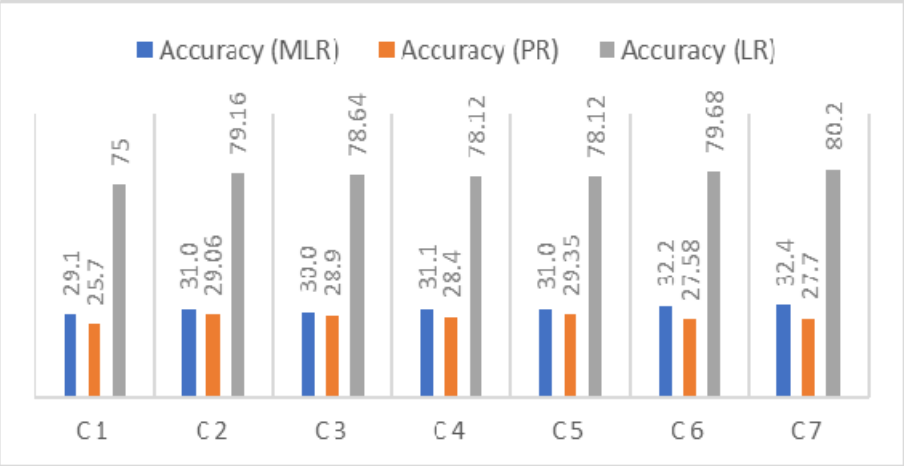


Fig. 8. Accuracy of CPMS for ordered combination of parameters

A comparative analysis of MLR, PR and LR for ordered combination of parameters in CPMS is shown in Table 5. A significant percentage increase of accuracy is identified when Logistic Regression is used instead of Multiple Linear Regression and Polynomial Regression.

Table 5. Comparative Analysis of MLR, PR and LR for ordered CPMS

CPMS	Accuracy (MLR)	Accuracy (PR)	Accuracy (LR)	Diff.(LR-MLR)	Diff.(LR-PR)	%Inc wrt MLR
C1	29.1	25.7	75	45.89	49.3	157.6
C2	31.0	29.06	79.16	48.13	50.1	155.1
C3	30.0	28.9	78.64	48.69	49.74	162.6
C4	31.1	28.4	78.12	47.06	49.72	151.5
C5	31.0	29.35	78.12	47.15	48.77	152.2
C6	32.2	27.58	79.68	47.48	52.1	147.5
C7	32.4	27.7	80.2	47.81	52.5	147.6

5. Conclusion

Diabetes stands as one of the most pervasive ailments afflicting people across the globe. This complex condition manifests through a range of parameters, often entailing severe and even life-threatening health implications. Swift identification of these parameters at an early stage holds the potential to preclude the emergence of severe health issues, offering substantial benefits to individuals. The current research is dedicated to a comprehensive parametric analysis of Patient Medical Statistics (PMS) and Combination of Patients Medical Statistics (CPMS) through the lens of linear, multi-linear, polynomial, and logistic regression models. The efficacy of these models is assessed using RMSE and R2 metrics, unveiling the logistic regression model's superior accuracy, attributed to its inherently discrete nature. Additionally, the study reveals that orchestrating CPMS in a systematic sequence according to their impact on diabetes detection yields heightened accuracy rates. Looking ahead, the study's implications suggest avenues for future exploration, such as broadening the scope of analyzed parameters for disease detection enhancement

References

[1] K. Gopinath, R. Jayakumararaj and M. Karthikeyan, "DAPD: A Knowledgebase for Diabetes Associated Proteins," in IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 12, no. 3, pp. 604-610, 1 May-June 2015, doi: 10.1109/TCBB.2014.2359442.

[2] C. Owens, H. Zisser, L. Jovanovic, B. Srinivasan, D. Bonvin and F. J. Doyle, "Run-to-run control of blood glucose concentrations for people with type 1 diabetes mellitus," in IEEE Transactions on Biomedical Engineering, vol. 53, no. 6, pp. 996-1005, June 2006, doi: 10.1109/TBME.2006.872818.

[3] M. E. Wilinska, L. J. Chassin, H. C. Schaller, L. Schaupp, T. R. Pieber and R. Hovorka, "Insulin kinetics in type-1 diabetes: continuous and bolus delivery of rapid acting insulin," in IEEE Transactions on Biomedical Engineering, vol. 52, no. 1, pp. 3-12, Jan. 2005, doi: 10.1109/TBME.2004.839639.

[4] S. Rahaman, "Diabetes diagnosis decision support system based on symptoms, signs and risk factor using special computational algorithm by rule base," 2012 15th International Conference on Computer and Information Technology (ICCIT), Chittagong, Bangladesh, 2012, pp. 65-71, doi: 10.1109/ICCITechn.2012.6509796.

[5] E. I. Georga et al., "Multivariate Prediction of Subcutaneous Glucose Concentration in Type 1 Diabetes Patients Based on Support Vector Regression," in IEEE Journal of Biomedical and Health Informatics, vol. 17, no. 1, pp. 71-81, Jan. 2013, doi: 10.1109/TITB.2012.2219876.

[6] C. Owens, H. Zisser, L. Jovanovic, B. Srinivasan, D. Bonvin and F. J. Doyle, "Run-to-run control of blood glucose concentrations for people with type 1 diabetes mellitus," in IEEE Transactions on Biomedical Engineering, vol. 53, no. 6, pp. 996-1005, June 2006, doi: 10.1109/TBME.2006.872818.

[7] B. J. Lee and J. Y. Kim, "Identification of Type 2 Diabetes Risk Factors Using Phenotypes Consisting of Anthropometry and Triglycerides based on Machine Learning," in IEEE Journal of Biomedical and Health Informatics, vol. 20, no. 1, pp. 39-46, Jan. 2016, doi: 10.1109/JBHI.2015.2396520.

[8] M. Cracchiolo et al., "Decoding Neural Metabolic Markers From the Carotid Sinus Nerve in a Type 2 Diabetes Model," in IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 27, no. 10, pp. 2034-2043, Oct. 2019, doi: 10.1109/TNSRE.2019.2942398.

[9] P. Colmegna et al., "Evaluation of a Web-Based Simulation Tool for Self-Management Support in Type 1 Diabetes: A Pilot Study," in IEEE Journal of Biomedical and Health Informatics, vol. 27, no. 1, pp. 515-525, Jan. 2023, doi: 10.1109/JBHI.2022.3209090.

[10] H. G. Clausen et al., "A New Stochastic Approach for Modeling Glycemic Disturbances in Type 2 Diabetes," in IEEE Transactions on Biomedical Engineering, vol. 68, no. 10, pp. 3161-3172, Oct. 2021, doi: 10.1109/TBME.2021.3074868.

[11] https://www.who.int/news-room/fact-sheets/detail/diabetes, [Date of Access: 06/04/2023]

[12] Mehan V., "Exploring the Future Jobs, Working Experience, Ethical Issues and Skills from Artificial Intelligence", International Journal of Innovative Science and Research Technology, Volume 8 , Issue 9 , Sept. - 2023.

[13] S. D. Patek et al., "Modular Closed-Loop Control of Diabetes," in IEEE Transactions on Biomedical Engineering, vol. 59, no. 11, pp. 2986-2999, Nov. 2012, doi: 10.1109/TBME.2012.2192930.

[14] W. Luangruangrong, A. Rodtook and S. Chimmanee, "Study of Type 2 diabetes risk factors using neural network for Thai people and tuning neural network parameters," 2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Seoul, Korea (South), 2012, pp. 991-996, doi: 10.1109/ICSMC.2012.6377858.

[15] C. Zhao and C. Yu, "Rapid Model Identification for Online Subcutaneous Glucose Concentration Prediction for New Subjects With Type I Diabetes," in IEEE Transactions on Biomedical Engineering, vol. 62, no. 5, pp. 1333-1344, May 2015, doi: 10.1109/TBME.2014.2387293.

[16] H. Abbas, L. Alic, M. Rios, M. Abdul-Ghani and K. Qaraqe, "Predicting Diabetes in Healthy Population through Machine Learning," 2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS), Cordoba, Spain, 2019, pp. 567-570, doi: 10.1109/CBMS.2019.00117.

[17] X. Zhang, J. Deng and R. Su, "The EM algorithm for a linear regression model with application to a diabetes data," 2016 International Conference on Progress in Informatics and Computing (PIC), Shanghai, China, 2016, pp. 114-118, doi: 10.1109/PIC.2016.7949477.

[18] Y. Zou, X. Gong, P. Miao and Y. Liu, "Using TensorFlow to Establish multivariable linear regression model to Predict Gestational Diabetes," 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), Chongqing, China, 2020, pp. 1695-1698, doi: 10.1109/ITNEC48623.2020.9084664.

[19] K. U. V. R. Teja, B. P. V. Reddy, L. P. A, H. Y. Patil and P. C. T, "Prediction of Diabetes at Early Stage with Supplementary Polynomial Features," 2021 Smart Technologies, Communication and Robotics (STCR), Sathyamangalam, India, 2021, pp. 1-5, doi: 10.1109/STCR51658.2021.9588849.

[20] A. Venkatesh. and M. S. Saravanan., "An Efficient Method for Predicting Linear Regression with Polynomial Regression," 2022 3rd International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 2022, pp. 1603-1606, doi: 10.1109/ICOSEC54921.2022.9952049.

[21] G. Oraz and X. Luo, "County-level geographic distributions of diabetes in relation to multiple factors in the united states," 2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI), Las Vegas, NV, USA, 2018, pp. 279-282, doi: 10.1109/BHI.2018.8333423.

[22] F. Simone, F. Andrea, S. Giovanni, P. Gianluigi and D. F. Simone, "Linear Model Identification for Personalized Prediction and Control in Diabetes," in IEEE Transactions on Biomedical Engineering, vol. 69, no. 2, pp. 558-568, Feb. 2022, doi: 10.1109/TBME.2021.3101589.
[23] V. Ganesh, J. Kolluri and K. V. Kumar, "Diabetes Prediction using Logistic Regression and Feature Normalization," 2021 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES), Chennai, India, 2021, pp. 1-6, doi: 10.1109/ICSES52305.2021.9633773.
[24] L. Lei, "Prediction of Score of Diabetes Progression Index Based on Logistic Regression Algorithm," 2020 International Conference on Virtual Reality and Intelligent Systems (ICVRIS), Zhangjiajie, China, 2020, pp. 954-956, doi: 10.1109/ICVRIS51417.2020.00232.
[25] A. Mangal and V. Jain, "Performance analysis of machine learning models for prediction of diabetes," 2022 2nd International Conference on Innovative Sustainable Computational Technologies (CISCT), Dehradun, India, 2022, pp. 1-4, doi: 10.1109/CISCT55310.2022.10046630.
[26] M. F. Faruque, Asaduzzaman and I. H. Sarker, "Performance Analysis of Machine Learning Techniques to Predict Diabetes Mellitus," 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), Cox'sBazar, Bangladesh, 2019, pp. 1-4, doi: 10.1109/ECACE.2019.8679365.
[27] T. H. Nasution and L. A. Harahap, "Predict the Percentage Error of LM35 Temperature Sensor Readings using Simple Linear Regression Analysis," 2020 4rd International Conference on Electrical, Telecommunication and Computer Engineering (ELTICOM), Medan, Indonesia, 2020, pp. 242-245, doi: 10.1109/ELTICOM50775.2020.9230472.
[28] C. -H. Wu, J. -B. Li and T. -Y. Chang, "SLinRA2S: A Simple Linear Regression Analysis Assisting System," 2013 IEEE 10th International Conference on e-Business Engineering, Coventry, UK, 2013, pp. 219-223, doi: 10.1109/ICEBE.2013.33.
[29] P. Wang, R. Ge, X. Xiao, M. Zhou and F. Zhou, "hMuLab: A Biomedical Hybrid MUlti-LABel Classifier Based on Multiple Linear Regression," in IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 14, no. 5, pp. 1173-1180, 1 Sept.-Oct. 2017, doi: 10.1109/TCBB.2016.2603507.
[30] H. T. Hoc, R. Silhavy, Z. Prokopova and P. Silhavy, "Comparing Multiple Linear Regression, Deep Learning and Multiple Perceptron for Functional Points Estimation," in IEEE Access, vol. 10, pp. 112187-112198, 2022, doi: 10.1109/ACCESS.2022.3215987.
[31] Y. Wang, L. Li and C. Dang, "Calibrating Classification Probabilities with Shape-Restricted Polynomial Regression," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 41, no. 8, pp. 1813-1827, 1 Aug. 2019, doi: 10.1109/TPAMI.2019.2895794.
[32] A. Venkatesh. and M. S. Saravanan., "An Efficient Method for Predicting Linear Regression with Polynomial Regression," 2022 3rd International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 2022, pp. 1603-1606, doi: 10.1109/ICOSEC54921.2022.9952049.
[33] K. He and C. He, "Housing Price Analysis Using Linear Regression and Logistic Regression: A Comprehensive Explanation Using Melbourne Real Estate Data," 2021 IEEE International Conference on Computing (ICOCO), Kuala Lumpur, Malaysia, 2021, pp. 241-246, doi: 10.1109/ICOCO53166.2021.9673533.
[34] J. Nie, J. Fang and Y. Zhao, "Cow Health Prediction Method Based on Logistic Regression and Decision Tree," 2022 34th Chinese Control and Decision Conference (CCDC), Hefei, China, 2022, pp. 3712-3717, doi: 10.1109/CCDC55256.2022.10033946.
[35] https://www.kaggle.com/datasets/mathchi/diabetes-data-set, Date of Access: 28-04-2023
[36] K. Rajesh and M. S. Saravanan, "Prediction of Customer Spending Score for the Shopping Mall using Gaussian Mixture Model comparing with Linear Spline Regression Algorithm to reduce Root Mean Square Error," 2022 6th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2022, pp. 335-341, doi: 10.1109/ICICCS53718.2022.9788162.
[37] Gregory Y.H. Lip, Ken Haguenoer, Christophe Saint-Etienne, Laurent Fauchier, "Relationship of the SAMe-TT2 R2 Score to Poor-Quality Anticoagulation, Stroke, Clinically Relevant Bleeding, and Mortality in Patients With Atrial Fibrillation," Chest, Volume 146, Issue 3, 2014, Pages 719-726, ISSN 0012-3692, https://doi.org/10.1378/chest.13-2976.

Conflict of Interest: The author has no conflicts of interest to declare.

Aut**hor Profile**

Prof. (Dr.) Vineet Mehan, received the B.Tech. degree from Kurukshetra University, M.E. degree from NITTTR Chandigarh and Ph.D. degree from NIT Jalandhar. He is currently a Professor with the Department of Artificial Intelligence and Machine Learning, Computer Science and Engineering, Chandigarh University, Punjab, India, since Jun. 2023. He has published over 50 papers in peer-reviewed international journals and conferences. Her research interests include Machine Learning, AI, Bio-informatics and DIP..